



THE UNIVERSITY *of* EDINBURGH

## Edinburgh Research Explorer

### **DlaaS: Data-Intensive workflows as a service - Enabling easy composition and deployment of data-intensive workflows on Virtual Research Environments**

#### **Citation for published version:**

Filgueira Vicente, R, Ferreira da Silva, R, Deelman, E & Atkinson, M 2016, DlaaS: Data-Intensive workflows as a service - Enabling easy composition and deployment of data-intensive workflows on Virtual Research Environments. in *American Geoscience Union Fall Meeting 2016*. American Geoscience Union Fall Meeting 2016, San Francisco , California, United States, 12/12/16.  
<<https://agu.confex.com/agu/fm16/meetingapp.cgi/Paper/156232>>

#### **Link:**

[Link to publication record in Edinburgh Research Explorer](#)

#### **Document Version:**

Peer reviewed version

#### **Published In:**

American Geoscience Union Fall Meeting 2016

#### **General rights**

Copyright for the publications made accessible via the Edinburgh Research Explorer is retained by the author(s) and / or other copyright owners and it is a condition of accessing these publications that users recognise and abide by the legal requirements associated with these rights.

#### **Take down policy**

The University of Edinburgh has made every reasonable effort to ensure that Edinburgh Research Explorer content complies with UK legislation. If you believe that the public display of this file breaches copyright please contact [openaccess@ed.ac.uk](mailto:openaccess@ed.ac.uk) providing details, and we will remove access to the work immediately and investigate your claim.



## ABSTRACT SUBMISSION CONTROL PANEL

### Abstract ID#:

156232

Date Started: 2016-08-02

06:17:37

Last Modified: now()

Primary Section/Focus

Group: IN

### 7 Steps to submit an Abstract:

- ✓ 1. [Session](#)
- ✓ 2. [Abstract Details](#)
- ✓ 3. [Author\(s\)](#)
- ✓ 4. [Abstract Text](#)
- ✓ 5. [Special Programs](#)
- ✓ 6. [Payment](#)
- ✓ 7. [Submit Abstract](#)

### Options:

[Abstract Tutorial](#)

### Sign out

[-View Submission](#)[-Withdraw](#)[-Transfer](#)[-Letter of Invitation](#)[-Abstracts Viewer](#)[-User Portal](#)

### Help:

[Report a Technical Issue](#)

## Your abstract submission has been received

Click [HERE](#) to print this page now.

**You have submitted the following abstract to 2016 AGU Fall Meeting. Receipt of this notice does not guarantee that your submission was complete or free of errors.**

### **DlaaS: Data-Intensive workflows as a service - Enabling easy composition and deployment of data-intensive workflows on Virtual Research Environments**

**Rosa Filgueira<sup>1</sup>**, Rafael Ferreira da Silva<sup>2</sup>, Ewa Deelman<sup>2</sup> and Malcolm Atkinson<sup>1</sup>, (1)University of Edinburgh, School of Informatics, Edinburgh, United Kingdom, (2)USC Information Sciences Inst., Marina Del Rey, CA, United States

### Abstract Text:

We present the *Data-Intensive workflows as a Service (DlaaS)* model for enabling easy data-intensive workflow composition and deployment on clouds using containers. *DlaaS* model backbone is *Asterism*, an integrated solution for running data-intensive stream-based applications on heterogeneous systems, which combines the benefits of *dispel4py* with *Pegasus* workflow systems. The stream-based executions of an *Asterism* workflow are managed by *dispel4py*, while the data movement between different e-Infrastructures, and the coordination of the application execution are automatically managed by *Pegasus*.

*DlaaS* combines *Asterism* framework with *Docker* containers to provide an integrated, complete, easy-to-use, portable approach to run data-intensive workflows on distributed platforms. Three containers integrate the *DlaaS* model: a *Pegasus* node, and an MPI and an Apache Storm clusters. Container images are described as *Dockerfiles* (available online at [http://github.com/dispel4py/pegasus\\_dispel4py](http://github.com/dispel4py/pegasus_dispel4py)), linked to *Docker Hub* for providing continuous integration (automated image builds), and image storing and sharing. In this model, all required software (workflow systems and execution engines) for running scientific applications are packed into the containers, which significantly reduces the effort (and possible human errors) required by scientists or VRE administrators to build such systems. The most common use of *DlaaS* will be to act as a backend of VREs or Scientific Gateways to run data-intensive applications, deploying cloud resources upon request.

We have demonstrated the feasibility of *DlaaS* using the data-intensive *seismic ambient noise cross-correlation* application (Figure 1). The application preprocesses (*Phase1*) and cross-correlates (*Phase2*) traces from several seismic stations. The application is submitted via *Pegasus (Container1)*, and *Phase1* and *Phase2* are executed in the MPI (*Container2*) and Storm (*Container3*) clusters respectively. Although both phases could be executed within the same environment, this setup demonstrates the flexibility of *DlaaS* to run applications across e-Infrastructures.